1972

Freedom of Information Collection
Research and Information Services
Ontario Ministry of Finance, Frost Bldg North, Main fl

Does not Circulate

ONTARIO

DEPARTMENT OF TREASURY AND ECONOMICS

9KUW

PARLIAMENT BUILDINGS
TORONTO

OVERVIEW OF 1971 CENSUS OF POPULATION AND HOUSING

February 1972

Economic and Statistical Services Division
Department of Treasury and Economics

Ontario
Ministry of Finance
9KUW
JUL 0 9 2004
Information Services

Prepared by
Ontario Statistical Centre
with the
assistance of
Electronic Computing Branch
Department of Transportation and Communications

Ont. Ministry of Treasury, Economics & Intergovernmental Affairs
51/1651
MAR 27 1981
LIBRARY

February 3, 1972

HA
747.65
.093
1972
c.1
tor mai

[illegible]

INDEX

Page

Introduction. 2

Collection of data - Dorothy Fetterly 4

Data Content - Ed McCoy 15

Processing the data - Shashi Sharma 20

Organization of the data on different media - Dave Weeks. 32

Geographical Levels - Dyke MacInnes. 39

Organization of the data for retrieval - Tony Wellman. 53

Software for census retrieval - Bill Gray 67

Dissemination of data at the Provincial level - Tony Wellman 75

Digitized by the Internet Archive
in 2018 with funding from
Ontario Council of University Libraries

https://archive.org/details/overviewof1971ce00onta

INTRODUCTION

INTRODUCTION

The modern census originated in Canada in 1666 when Jean Talon took an official census to measure the increase in population of the colony since the founding of Quebec by Champlain in 1608.

Talon's enumeration, recording a total of 3215 persons included the name, age, sex, marital status and occupation of every person.

In 1871 a tabulation of all preceding censuses in what is now Canada was made covering a period of more than 200 years. Thus Canada became the only country in the world to have statistical records of its growth from its earliest days.

The increasing complexity of society and the increasing role of government, has created the need for more and more information about the population in order that both government and the private sector may plan constructively.

However, only because of the development of sophisticated processing techniques has it become feasible to collect and process the quantity of information collected on the 1971 census form.

In all its configurations and cross tabulations possible, the full potentialities of the information in policy making and planning are immense.

Today we hope to give you some feel of what the 1971 census is all about in order that you might explore these potentialities yourselves.

COLLECTION OF DATA

Dorothy Fetterly

Census 1971

Collection of Data, Sampling and Self-enumeration and Follow-up

Planning for the 1971 Census

- the assessment of user needs, the testing and evaluation programmes

Brief Outline of Census Framework

- dwelling defined, household defined

Self-enumeration

- forms 2A and 2B - the questionnaires were delivered to each household and completed by each family.
- role of the census representative or enumerator. The Visitation Record. The enumeration area.

Sampling

- how the sampling procedure was carried out in 1 of 3 households. The difference between the long and short questionnaires. Reasons for sampling. Sampling studies in the Census Evaluation Programme.

Census 1971
Collection of Data, Sampling, Self-enumeration and Follow-up

Planning for the 1971 Census

The 1971 Census had a more comprehensive advance testing and evaluation program than any previous census. A basic element in the planning was an investigation of users needs for census data as well as a programme to test various methods of collecting and processing the data.

In the assessment of the users' needs, the nature of demands following previous censuses was taken into consideration, but the main thrust of this phase of the planning was a formal liaison programme with federal, provincial and other groups, "other" including municipalities, universities, the business community, and social and religious organizations.

The second element or phase in the planning was the testing programme. Research into the quality of the 1961 Census data and studies of international experience of a similar nature especially in the U.S.A. identified several important sources of error to which census statistics are subject. Examples are inadvertent or deliberate errors by respondents, misinterpretation of instructions or responses by enumerators and errors at the editing and processing stages. The studies indicated that by far the greatest reduction in error could be achieved if the role of the enumerator in the data collection process could be minimized, and it became an objective of the testing programme to bring this about. In the period between 1966 and 1969, several

variations of methods involving self-enumeration were tested and the merits of various field procedures were investigated. Tests were carried out in Ottawa, London, Toronto and four rural areas, and a Trial Census was held in September 1969.

In addition to determining the field procedures for 1971, the testing programme provided the basis for decisions on such topics as questionnaire content and design, the degree of sampling to be used, and certain processing operations related to the "self-enumeration" technique.

Brief Outline of Census Framework

Dwelling defined

It is the places where people live or may live - their dwellings - around which both the Census of Population and the Census of Housing are built. The people who live in these dwellings, both farm and non-farm, provide the universe for the Census of Population. The dwellings, whether occupied or not, are the building blocks for the Census of Housing. Units that were identified as dwellings were classified into two broad groups, private dwellings and collective dwellings. The bulk of the population will be found in the first of these groups.

Included as private dwellings were those in which a family or other small group of individuals could reside (e.g. a single house or an apartment). Collectives included those special units such as hotels, large lodging houses, hospitals, and other types of group quarters that could house fairly large numbers of unrelated persons.

Household defined

The identification of a household follows the identification of each occupied dwelling, since the group of persons occupying each of these units comprise a household. Therefore, the number of occupied dwellings and the number of households are the same, and there is a tendency to use these terms interchangeably. The size and type of household may vary and a household may consist of one person living alone, a family, several unrelated persons living together in a private dwelling, or a larger group occupying a collective dwelling such as a hotel or hospital.

The Census is based on the principle of enumerating every person at his usual place of residence. Criteria based on usual or temporary residence were the factors that determined who should be included as members of each household. Basically, every person living in Canada on the census date (i.e., at midnight between May 31 and June 1) was to be counted as a member of the household where he resided.

Self-enumeration

The method of enumeration in the 1971 Census of Population and Housing was that of self-enumeration rather than census taking by a Census Representative or enumerator. In the past, census-takers or enumerators called on every household to ask certain questions of the head of the household and to record his answers. In the 1971 Census, this method was used only in remote parts of Canada where it was more difficult for the Head of the household to mail back his questionnaire or for a Census Representative to call back a second

time to pick it up. During the week before census day (June 1, 1971), Census Representatives called at each household to leave a questionnaire and an Instruction Booklet. Each family was able to study the questionnaire to ensure that by June 1 they could complete every applicable question with accurate information. Most of the questions were to be answered by filling in a small circle with a black lead pencil. The completed questionnaire was returned in one of two ways.

Families living in urban areas were asked to mail their questionnaires back in the addressed, postage paid envelope provided, while households in all other areas kept their questionnaires until the Census Representative called to pick them up. These procedures were adopted for more than 97 percent of the population in 1971. The remainder, including persons residing in remote northern areas, institutions, and military barracks were covered by traditional canvasser methods. Although the adoption of self-enumeration techniques reduced the role of the interviewer, he was required to ensure that every household in the area assigned to him was included and that the questionnaires returned were adequately completed. This follow-up was by telephone or by personal contact to obtain the necessary information. In the pick-up area, this follow-up was completed at the time the documents were collected.

In the actual enumeration process, the federal census respects the existing boundaries, i.e. the boundaries of electoral districts (ED's) and municipal boundaries defined for administrative purposes by the provinces. The electoral district serves as a backdrop to the creation

of enumeration areas (EA's) which constitute the basic unit in the collection of census data.

In each Census the boundaries and the numbering of enumeration areas change since the essential factor in their delineation is population.

Very briefly, the EA usually represents the territory an enumerator can cover in the period assigned to him. Several criteria govern the delineation of enumeration areas:

- population: approximately 250 households or 100 farms;
- recognizable boundaries: waterway, railway, road;
- homogeneous rural or urban character;
- easy accessibility of each section of the territory;
- respect for the boundaries of other statistical or administrative units;

The basic document used by the enumerator (or Census Representative) was the Visitation Record. In the V.R., the enumerator listed all the households in the E.A. by an assigned household number, and every third number was indicated or marked to receive a long form to fill out. The Visitation Record listed such information as date of visit, the exact location of the household, the number of usual residents and the name of the household head. The above data was recorded under one of three classifications: private - type dwellings - occupied, private - type dwellings - vacant, and collective - type dwellings.

In order to gain the greatest accuracy possible, the information given on the questionnaires is checked against other government information, such as that given in the Labour Force Survey, the Job Vacancy Survey, Family Allowance records, CMHC records, Canadian Real Estate Association records and Income Tax Statistics.

The use of self-enumeration with follow-up in the 1971 Census is expected to represent a substantial improvement over previous censuses for several reasons:

(1) the enumerators' contribution to the response errors will be substantially diminished since the role of enumerators is greatly reduced.

(2) each adult member of the household is able to answer the census questions himself. In previous censuses, the housewife very often had to provide all the complex and detailed information to the enumerator concerning all members of her household.

(3) respondents are able to consult their records.

(4) the enumerator can review the census returns and collect the missing information.

Sampling

The 1971 Census was also characterized by the use of sampling at a level higher than had ever been used in the census before. Two questionnaires were used for the Population and Housing Censuses.

Form 2A, the short questionnaire, was left at two households, and Form 2B, the "long" questionnaire or the one containing the sample questions in addition to the questions included on the short form, was left at the third household. Only one round of visits was made to a given area and a random start was made to prevent any sampling bias. Some enumerators left long forms at the 1st, 4th, 7th households, others at the 2nd, 5th, 8th and the rest at the 3rd, 6th and 9th in their respective areas. As a result of this procedure, two-thirds of all households were asked fewer questions than in 1961 (averaging 9 per person as compared to 20 in 1961) and only one-third were asked significantly more questions (averaging 48 per person as compared to 36 in 1961).

The short questionnaire received by two-thirds of all households in Canada provides DBS with the basic population and housing data. Households report the name, birth date, family relationship, sex, mother tongue and marital status of each household member and some questions concerning their housing. The long questionnaire included these same questions plus others to provide more information to get a more detailed picture of the Canadian people: place of birth, ethnic origins, education, war service, employment and income details for each member of the household.

It was necessary for complete coverage on a number of basic characteristics because a complete count is required by the B.N.A. Act and bench mark statistics are necessary at certain intervals to

supplement population projections. However, the decision on the size of the sampling fraction was mainly the result of balancing the savings due to sampling against the loss in accuracy of the census data. Sampling enables considerable savings in terms of cost, timeliness and burden on the public. There are cost savings in both field operations and processing operations, at both the regional level and head office level. There is a reduction of total time to publication of census data because it reduces the time taken to process the census forms and produce the statistical data. Lastly, sampling reduces the burden on the public in completing questionnaires and providing the information required in the Census.

The details of the sampling procedures and some estimates of the size of errors associated with varying sampling fractions are provided in DBS papers. Neither a full count nor a sample census is free from errors, and, therefore, it is the relative magnitude of errors which determines the method to be used. In the Census Evaluation Programme, two projects have a direct bearing in the sampling procedures: the Sampling Ratio Project and the Sampling Bias Project. The Sampling Ratio Project is basically a comparison of cost, timing and accuracy for differing sampling ratios. As opposed to estimates based on previous Census or Trial Census data, this study will use the actual data (costs, errors, etc.) from the 1971 Census. The Sampling Bias Project attempts to estimate the amount of enumeration bias introduced at the drop-off stage due to the enumerator leaving long forms at the wrong households.

Summary

This paper has been an attempt to outline some of the procedures in the 1971 Census for the collection of the data, particularly those relating to the introduction of the self-enumeration techniques and the sampling operation. The 1971 Census is expected to provide more reliable data and more available data than achieved in 1961.

DATA CONTENT

Ed McCoy

Data Content

The content of the 1971 Census questionnaire was developed by an evolutionary process drawing upon Statistics Canada experiences of past Censuses and public demand. The decisions as to the inclusion or exclusion of specific questions from the 1971 Census were based on four criteria:

(1) The value of the question for the administration of government and/or to a relatively broad group of users.

(2) The ease (or difficulty) in obtaining reliable or factual data.

(3) Logistics of questionnaire space.

(4) A practical consideration of respondent tolerance.

The increasing importance attached to the Census of Population and Households as a unique source of socio-economic data is reflected by the development of:

(1) Additional depth to selected questions carried over from previous censuses, eg. education.

(2) Inclusion of new questions, eg. question 25 which relates to vocational training courses, question 27 which relates to moves in the previous 5 year period.

(3) A Census system which would be capable of using new technology.

(4) Extensive programs designed to improve the quality of the census data.

(5) More extensive consultation in planning the census questionnaire and format, to ensure the relevance of the data to the user.

The inclusion of new content in the 1971 Census questionnaire resulted from the increased need for statistical data vital in the general research and policy planning spheres. In the past regional analysis has been hampered by the lack of relevant statistical data. The evaluation of existing social programmes as well as the identification of the need for new ones is dependent on the timely availability of adequate data.

The Census of Population and Households may be grouped into four categories:

(1) Demographic and Social Characteristics

(2) Economic Characteristics

(3) Housing and Family Characteristics

(4) Administrative

The fourth category has no statistical significance.

In the past there has been a tendency to regard the census as a inventory procedure which counted the population and items directly related to the state of this population. Questions such as age, sex, marital status, religion, language and ethnic group have been included on every Census ever taken in Canada. While the census, is an inventory it is not restricted to this one function. Census data has proved invaluable as a unique source of statistical information in many areas in the decision making process of government. In many ways it represents the only source currently available to those concerned with planning and policy formulation. The Census is not only capable of determining the current state of the socio economic conditions but assists in helping in decisions as to where current policies may lead.

Our posture at this time is not to discuss each question included on the questionnaire but rather to explore the use of the response variables in terms of their information generating potential. The census data base in the past has been successfully exploited in the development of various profile series. The profile series products can cover a wide range both spatially and functionally.

Spatial arrangements could involve the following:

(1) Towns
(2) Census Tracts
(3) Census Division
(4) School districts
(5) Counties
(6) Economic regions
(7) Health districts
(8) Others which could be derived through a particular aggregation of EA's

Functional arrangements might consist of the following:

(1) Labour force
(2) Elderly
(3) Students
(4) Age distributions
(5) Income distributions
(6) Urban/rural resident
(7) Immigrants

As an example a housing profile which is in some respects a characterization of housing in Ontario; in specific urban centres as well as counties and municipalities could derive statistics such as the following:

(1) Average number of rooms in a housing unit

(2) Kitchen facilities

(3) Mobile homes

(4) Occupancy/vacancy status - the housing units vacant might be further divided into:

  (a) Those for sale or rent

  (b) Those not for sale or rent

(5) Quality of housing

(6) Value of housing, etc.

When profiles are available for well defined geographic units it is possible to draw comparisons and identify changes in the decade 1961 - 1971. Significant changes would alert planners to further explore other characteristics of the population in an attempt to explain the new patterns. One characteristic which might be suggested would be out migration. A portion of the labour force who because of economic conditions in the region may have moved to other areas of increased economic activity. Census data has value in directing attention to rather precise areas.

A profile of education could be derived in similar fashion and while no health related questions appear on the questionnaire the use of statistics derived from administrative records could be complemented by socio economics data from the census to provide a new source of statistical information.

PROCESSING OF THE 1971 CENSUS DATA

Shashi Sharma

Processing of the 1971 Census Data

1. Introduction

After data collection and quality control of enumerator assignments, the data is sent for processing. This includes editing, coding, imputing missing information, geographical organization and storage of data files, retrieval of information, synthesizing and tabulation of data. Each of these procedures could introduce errors in the data. Appropriate quality control procedures have thus been designed to evaluate and control outgoing quality from each of the phases of operation.

The processing consists mainly of two elements, namely clerical and computer processing. The clerical operations are further divided into two stages: Regional office processing and Head Office processing. Except for a very few operations, all pre-computer processing is done in the eight regional offices, while the manual processing in Head office mainly complements the computer processing.

2. Regional Office Processing

The coding of population and housing documents is done at the Regional offices. Two independent groups of coders do the coding of the documents. Each group codes only that portion of the document assigned to it.

(a) Economic Characteristics Coders: This group codes three items, namely occupation, industry and place of work. The occupations are coded to 4-digit level using a special edition of the newly developed Canadian Classification and Dictionary of Occupations (CCDO). Industry are

coded to 3 digits. The classification used in this coding is the Standard Industrial Classification Manual (SIC) 1970 edition. Place of work is a new topic in the Canadian census. In 1971 Census, place of work is coded to municipality level. After the main census processing, including computer processing, is complete, and if resources permit, Statistics Canada will undertake to code place of work on request to finer geographical areas such as census tracts.

(b) General Coders: This group codes mother tongue, citizenship, religion, vocational training, migration, income. Vocational training is a new topic added since 1961 to the questionnaire. The details of the vocational courses are coded to a 2-digit level comprising 72 distinct categories.

3. Head Office Processing

At Head Office the questionnaires are sent for microfilming and computer processing. The bulk of the Head Office manual operations are geared to preparing questionnaires for microfilming, transcribing damaged questionnaires, checking outputs from computers, and correcting errors. Through these operations it is possible to maintain control over the quality of data going into the computer and to ensure that the computerized editing and imputation procedures have not introduced distortions to the data.

4. Fosdic Operation

Transferring data from the Population and Housing Questionnaires to suitable computer readable magnetic tapes is accomplished by the use of "FOSDIC 70" equipment.

FOSDIC, which stands for Film Optical Sensing Device for Input to Computers, was developed jointly by the U.S. Bureau of Census and the National Bureau of Standards, Washington, D.C. An earlier model "FOSDIC 60" was used to process the 1960 U.S. Census. An upgraded and transistorized version of this machine, "FOSDIC 70" is being used by the U.S. Bureau to process the 1970 U.S. Census.

The questionnaires are microfilmed using a special 16mm. datapak thin film. The film is passed through FOSDIC, which scans the images of the questionnaires and records the answers on a magnetic tape. In effect, FOSDIC is a special purpose computer which has the capability of optical scanning of microfilm.

The FOSDIC operation is done in two stages. First, pages 2 and 3 of both Short (2A) and Long (2B) Forms are microfilmed and read by FOSDIC. The resulting tapes are subjected to a series of computer programmes for editing and imputation. Rules for imputing missing information have been carefully designed to take advantage of correlations between variables and to ensure that imputations do not lead to inconsistencies within households and within individual records.

The second stage consists of the processing of the long forms. The microfilming, FOSDIC, editing and imputation procedures are similar to those used for Stage I processing, except that missing data for occupation, industry, place of work and migration is not imputed, but left as unspecified.

5. The Editing and Imputation Programme

The 1971 Census editing procedures begin when the document information, organized in EA (Enumeration Area) batches, is transferred onto computer tape via a microfilming and FOSDIC step. The FOSDIC software, itself, has some minor editing features such as the selection of one entry when more than one response is received for a question.

To ensure that the information on tape adequately represents the information on the documents, a monitoring programme, called DIARY I, was appended to the FOSDIC operation. Prior to the reading of each EA batch of documents, FOSDIC is provided with an EA Breaker Sheet which contains, besides the geographic code, other pertinent data such as the population count, the document count and the vacant dwelling count. If the FOSDIC derived information does not match that of the Breaker Sheet, within very narrow tolerance limits, warning flags are produced. In addition, if the nonresponse for any question is above a pre-specified tolerance, that information is also printed by DIARY I.

DIARY I ensures that no EA is passed to the editing programme proper, unless the information on tape, for that EA, is an accurate representation of the information on the documents.

Having passed the DIARY I hurdle, the EA information is then ready to be edited. The information on tape is raw data and raw data inevitably contains errors. There may be blanks where data should appear, because of non-response on the part of the householder, from poor coding,

substandard microfilming, or FOSDIC failure which somehow passed DIARY I without detection. In addition, there may be inaccurate or inconsistent data because of respondent error, carelessness or ignorance, or because of inaccurate coding of write-in entries. Before this information can be used as input to subsequent weighting and tabulation systems, it is necessary that it be in the form required by these systems. It is the purpose of the Edit and Imputation Programme to clean up this raw information as far as possible by:

(1) filling all blank entries that should be filled;

(2) ensuring that each field contains valid data (i.e., within a certain range);

(3) ensuring that data within each individual record is consistent;

(4) carrying out various other operations in preparation for subsequent weighting and tabulation programmes.

Note that inaccurate data cannot be recognized unless they are also invalid or inconsistent.

(a) Editing Stage

The first step in editing is to check for valid codes. If a code is not valid, a special code (meaning invalid) is inserted. Also at this stage households may be re-structured if they do not satisfy the definition of a household, and persons within households may be re-ordered into a standard order (based on relationship to head). Cancellations are also effected at this stage.

The next stage is to check each record for certain internal consistencies. These include: .

Marital status - age, for each person
Relationship - sex, for each person
Number of dwellings - type of dwelling
Income - occupation, industry, age

These checks are limited to a few basic consistencies, and the majority of valid data will be left unchanged however unlikely they may appear. Thus, the impossibilities are eliminated, the improbabilities may remain. A system of priorities has been set up to establish which code(s) should be changed in records where inconsistencies are found.

(b) Imputation Stage

The imputation process involves the assignment of a valid and consistent code in cases where

(1) an original entry failed a consistency edit;
(2) an original entry was invalid;
(3) there was no original entry.

As far as possible imputation is carried out deterministically based on other data in the record. For example, if a wife's age is missing and the husband's age is known, a rule is given for imputing the wife's age from that of her husband. However, there are some cases where there is no logical basis for imputation either because there are no suitable alternative variables on which to base imputation, or because data are missing for these other variables. In these cases imputation is achieved by a random hotdeck method.

The hotdeck method required pre-defined imputation groups (or hotdecks) for each variable or group of variables on the file. Every time a variable is read from a record belonging to a particular imputation group (which implies that the variables defining the imputation group must be valid on the record), the value of that variable replaces the previous value of that variable in the appropriate hotdeck. Then, when a record requiring an imputed value for a particular variable is read, the imputed value is obtained by reading the value currently stored in the appropriate hotdeck (which is the value contained on the last record falling in the same imputation group).

Under the assumption that imputations are required at random throughout the file, the hotdeck method preserves the mean and variance of the distribution of each variable within each imputation group.

When all editing and imputation has been completed, the computer files should be clean, consistent, complete and ready as input to the weighting and tabulation programmes.

After the editing and imputation procedure is completed, a record of the net change per question for each EA is maintained by the DIARY II programme. In a year or so, an analysis of the output from DIARY II showing the effects of the edits on the data will be released. At present it may be sufficient to say that the edits

may have a major effect on small-area statistics such as the data for enumeration areas, but their effect diminishes as the population of the study area increases.

6. Weighting Procedures

The next phase in processing of sample data is to determine the weighting factors for individuals and households. The weights are used to estimate population distributions from sample data. Each individual and each household in the sample carries an integral weight. These weights are used to compile the tables from sample data in which each individual (or household) is counted as many times as the corresponding weight.

The procedure used to determine the weights is known as "The Raking Ratio Estimation Procedure". This procedure takes the correlations between the sampled data and basic personal characteristics into account. It also reduces the variance of the weights. A separate set of weights employing similar procedures are calculated for the Agriculture Population linked file.

After the weights have been added to the sample file, a Master population file is available for the tabulation programme.

7. 1971 Evaluation Programme

The 1971 Census Evaluation programme consists of some 50 separate studies designed to determine the extent of undercoverage of census data, to assess the accuracy and reliability of the data and to measure the effectiveness of the procedures used to carry out the census.

A vital component of any census is the evaluation of its quality. Users need to know the degree of reliability of the results in terms of both coverage and content. Census officials analyse the quality of the data as well as evaluate the effectiveness of the many operations in order to plan for future censuses. In a period when advances in census technology are being made on unprecedented scales (e.g. self-enumeration by mail, document linkages, new data access and retrieval systems), an intensive evaluation programme is a major step in planning for economical censuses of high quality. Future censuses must meet the challenges of greatly expanded data needs of social and economic planners, and yet remain within budgets which governments will be prepared to finance. The most important of the coverage studies referred to above is the Reverse Record Check which will attempt to determine the amount of underenumeration and to identify those population sub-groups where underenumeration is highest.

In 1971, for the first time in Canada, most persons enumerated themselves and studies are required to evaluate the quality of the response in self-enumeration.

A Response Variance Study estimates the total error for different sizes of census estimates and for different characteristics or groups of characteristics at the regional level. These results will be published at the same time as the main body of census data. A long term research project is under taken by Statistics Canada which will give a more detailed breakdown of the components of the error, distinguishing between self-enumerated and enumerator-recorded responses and between Pick-up and Mail-back areas.

Another study to gauge the quality of response is the Initial Response Rate Study. It will identify the source of response, i.e. from the respondent or from follow-up or from computer imputation. Micro-comparisions of aggregates with previous census data and with records from other federal departments and agencies and micro-comparisons of census data for individuals with similar information from other sources will round out the evaluation of the consistency of census data. The most important of these matches is the match of census data with the Monthly Labour Force Survey data.

Extending the use of sampling in 1971 has created the need for studies evaluating the effect of sampling on census data. Attempts are being made to detect the causes of sampling bias and to measure the impact of sampling on time and costs.

New field procedures such as the Telephone Assistance Service, the procedures for resolving non-response in the field as well as processing procedures such as the Edit and Imputation programme, the coding of questionnaires, the Agriculture-Population linkage programme, etc. are being examined closely to determine their capacity for improving the efficiency of the census operation.

The geographical preparations for the census, such as the preparation of census tracts, the cartographic programme, the delineations of Census Metropolitan and urbanized areas and the Geocoding Programme for retrieval of small-area data are being evaluated to determine how well they have fulfilled their objective of providing the spatial framework for census data.

Other studies are being done to analyze the distribution of man-hours by function (Work Measurement Programme), evaluate the Quality Control procedures and provide a subjective appraisal of all aspects of the census from the training of personnel to the processing of the data.

Most of the evaluation studies will be carried out in the period 1971 to 1974 and their results will be documented and released as part of the Census Publication programme by Statistics Canada.

8. CONFIDENTIALITY

As yet there is no official policy on confidentiality. At present the report prepared by Census Working Group on Confidentiality Procedures is being 'considered' by the Census AD HOC Committee on confidentiality. This report considers the random rounding technique, which provides the strongest possible protection against every type of disclosure considered, and, of all methods considered, has the weakest effect upon the analytical meaning of the data. Furthermore, random rounding is the only technique considered which will permit the user of the data to aggregate published figures with confidence. The working group finding shows that, it is by far the cheapest and quickest technique to implement and the most desirable techniques on all criteria considered by the group. After final decision by the AD HOC Committee the confidentiality procedure could by discussed in detail.

ORGANIZATION OF THE DATA ON DIFFERENT MEDIA

Dave Weeks

ORGANIZATION OF THE 1971 CENSUS DATA ON DIFFERENT MEDIA

The 1971 Census data will exist in several different forms. Some will be held only by Statistics Canada and some by Ontario Statistical Centre. However, regardless of the media or location, O.S.C. will act as clearing-house for all census data requests from Ontario. The following forms of data storage will be used.

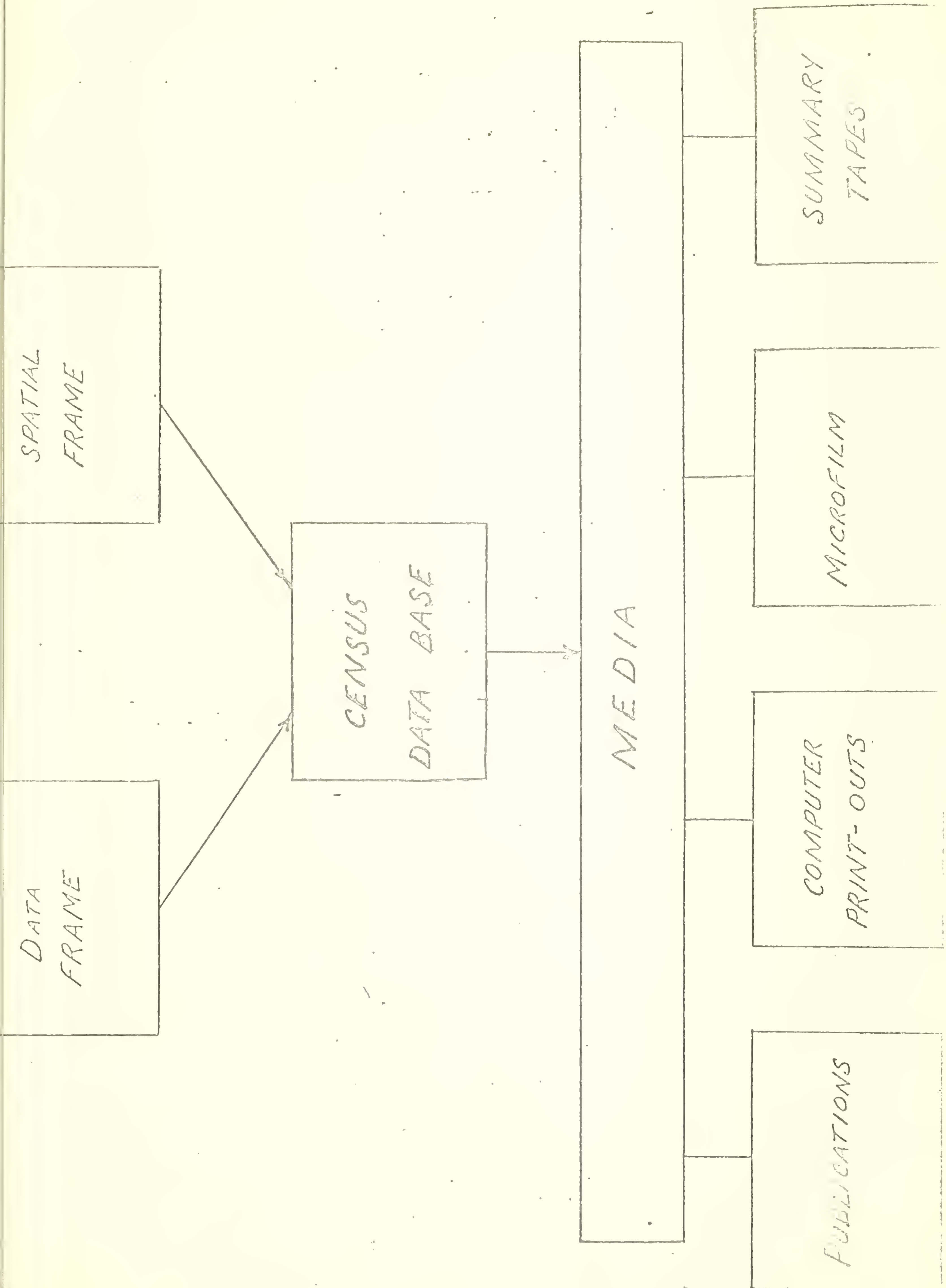
Medium	Description	Use
Master Files (on Tape)	Complete detail	Held by Statistics Canada
Geocoded Data Base (on Disk)	Complete detail	Held by S.C. to answer special requests, for example small area statistics
Production Summary Tapes	Complete cross tabulations aggregated by Enumeration Area	Held by S.C. to produce user summary tapes and publications
User Summary Tapes	Partial cross tabulations aggregated by various census areas (See page 34)	Held by O.S.C. used to produce tabulations by user-specified area
Computer Listings	Tabulations (some having restricted distribution because of confidentiality)	Held by O.S.C. on microfilm (except confidential data) for ready access for satisfying user requests
Microfilm	Same as for Computer Listings above, plus maps of census areas	Held by O.S.C.
Publications	Regular Census publication programme	Distributed by S.C. or Province upon request. Held by O.S.C. on microfilm for ready access and distribution of specific tables for users.

The User Summary Tapes, mentioned above, will exist in various levels depending on the size of the geographic unit as shown in the following table. As the size of the area increases, the amount of detail provided in the data cross tabulations also increases.

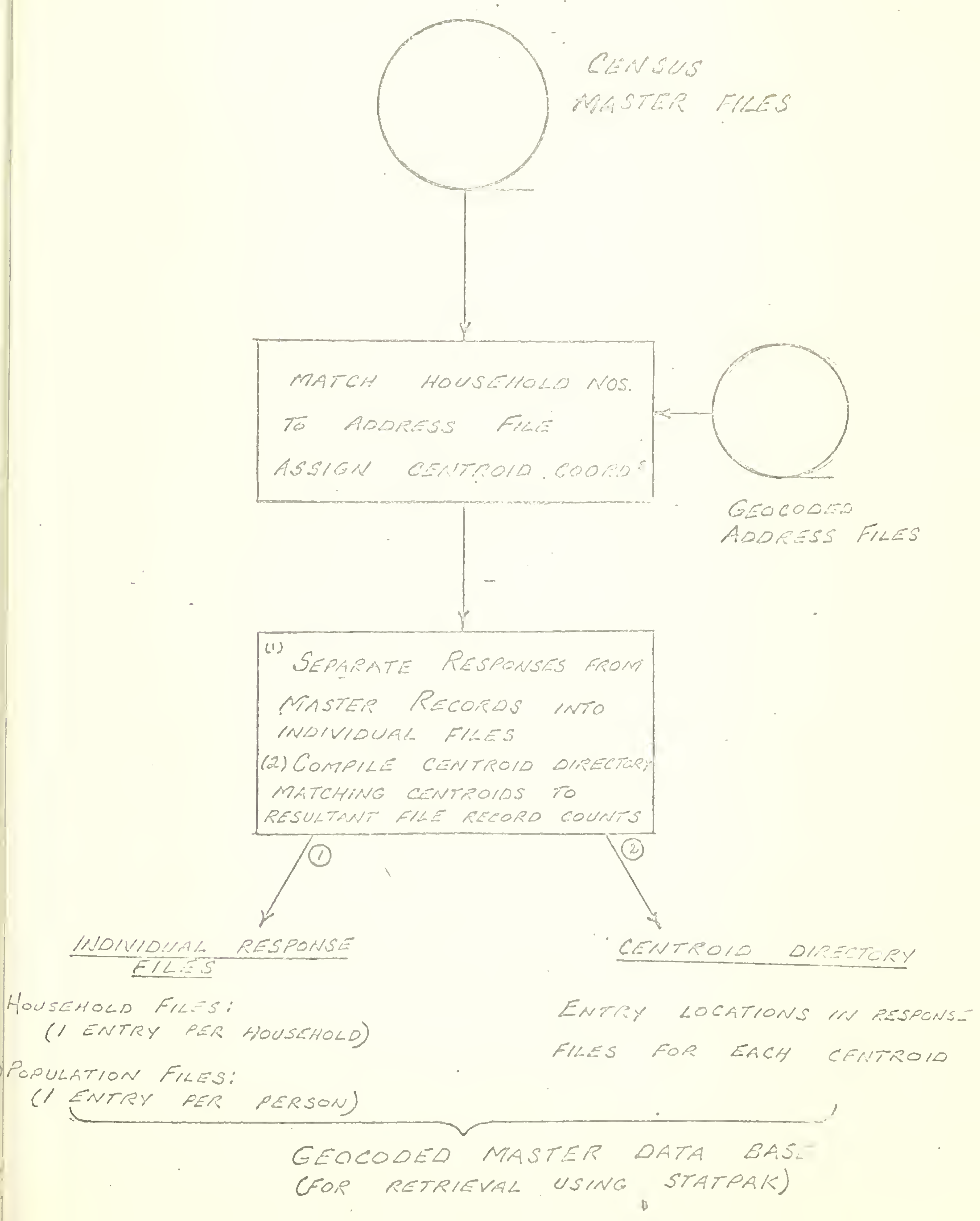
Summary Tape Level	Geographic Units		Data Characteristics
	Census Series	Municipal Series	
1	Enumeration Areas (EA)	All Municipalities	Totals for Individual Characteristics
2	Census Tract (CT) and Area Aggregate (AA)	Municipalities of 5000 population and above	Cross Tabulation of a few Characteristics
3	Census Division (CD) and County	Municipalities of 30,000 population and above	Cross Tabulation of many characteristics

Access to the census data will be facilitated by written documentation and maps. Data may be obtained for standard census areas such as Enumeration Areas or Counties. It may also be obtained for user specified areas such as school districts or traffic planning zones.

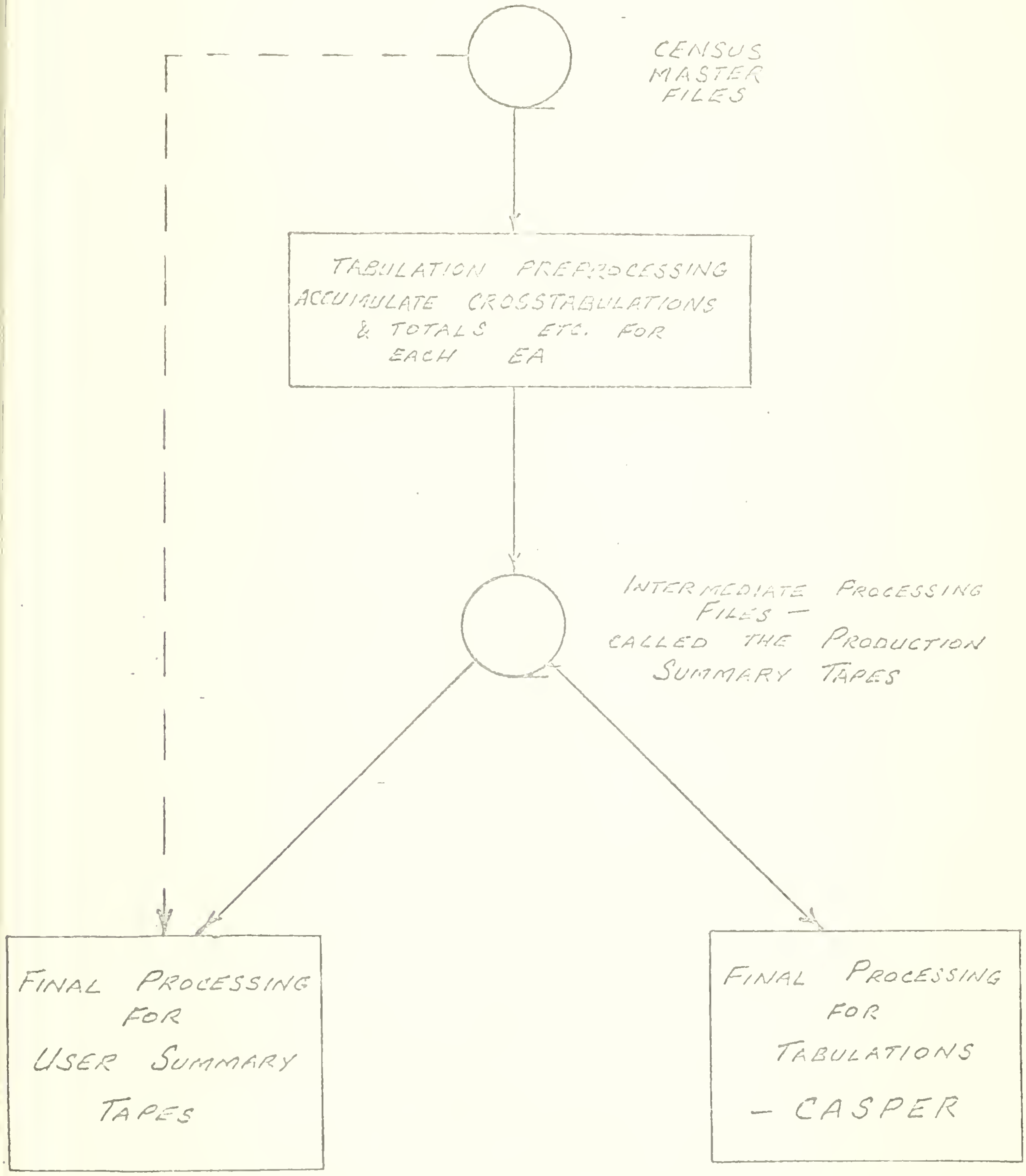
Listings of the tabulations in the programmed tabulation programme will be available at the workshop. Also, the latest available estimates of availability dates for tapes, etc. will be given.


SPATIAL FRAME
DATA FRAME
CENSUS DATA BASE
MEDIA
SUMMARY TAPES
MICROFILM
COMPUTER PRINT-OUTS
PUBLICATIONS

GEOCODING - GRDSR


CENSUS MASTER FILES
MATCH HOUSEHOLD NOS. TO ADDRESS FILE
ASSIGN CENTROID COORDS
GEOCODED ADDRESS FILES
(1) SEPARATE RESPONSES FROM MASTER RECORDS INTO INDIVIDUAL FILES
(2) COMPILE CENTROID DIRECTORY MATCHING CENTROIDS TO RESULTANT FILE RECORD COUNTS
①
②
INDIVIDUAL RESPONSE FILES
HOUSEHOLD FILES:
(1 ENTRY PER HOUSEHOLD)
POPULATION FILES:
(1 ENTRY PER PERSON)
CENTROID DIRECTORY
ENTRY LOCATIONS IN RESPONSE FILES FOR EACH CENTROID
GEOCODED MASTER DATA BASE
(FOR RETRIEVAL USING STATPAK)

PREPLANNED TABULATIONS (INCLUDING SUMMARY TAPES & MICROFILMS)


CENSUS MASTER FILES
TABULATION PREPROCESSING ACCUMULATE CROSSTABULATIONS & TOTALS ETC. FOR EACH EA
INTERMEDIATE PROCESSING FILES — CALLED THE PRODUCTION SUMMARY TAPES
FINAL PROCESSING FOR USER SUMMARY TAPES
FINAL PROCESSING FOR TABULATIONS — CASPER

THE CONTENT OF A 'SPECIAL' REQUEST

Whenever a request for data cannot be satisfied from one of the preplanned tabulations, which includes the user summary tapes, then it is designated as a 'special' request.

A special request, therefore, is one which requires access to one of the Main files held by Statistics Canada. Essentially there are three sets of main files:

(1) For General Retrievals and Tabulations requiring, for example, greater response detail than available from the user summary tapes -

- THE PRODUCTION SUMMARY TAPES (Intermediate files created as shown in the TABULATION processing)

(2) For Specific Area Retrievals and Tabulations requiring, for example, small area statistics -

- THE GEOCODED DATA BASE

(3) For General Retrievals and Tabulations requiring, for example, maximum response detail, or cross-correlations, not available from any summary tapes -

- THE CENSUS MASTER TAPES

The tools for these retrievals are all computer programs and program Packages:

for (1) the primary tool is the CASPER package
(2) GRDSR AND STATPAK
(3) Special programming and also the CASPER Package

GEOGRAPHICAL LEVELS

Dyke MacInnes

CENSUS GEOGRAPHICAL AREAS

Introduction:

Every request for data from the Census will require some reference to a geographical area. In the collection and publication phases of the decennial census, Statistics Canada recognizes, and in some cases creates specific geographical units. A brief review of these units will much improve our understanding of the levels at which census data is readily available.

Mr. Week's paper has mentioned that the amount of cross tabulated data varies with the size of the geographic unit which it represents. Table 1 shows all geographic levels.

TABLE 1

GEOGRAPHIC UNIT	SUMMARY TAPE LEVEL
Electoral District	3
County	3
Census Metropolitan Area	3
Census Agglomeration	2 or 3
Municipality	2 or 3
Area Aggregate	2
Census Tract	2
Census District	1
Enumeration Area	1

This discussion will proceed systematically from largest to smallest geographic areas and will also cover the different types and scales of maps available and *The Ontario Geographical Classification Manual for Production of Statistics*, which may be of some use in the 1971 Census Data Program.

Electoral District

Definition: An Electoral District is the area represented in the Federal Parliament by one member.

Location: The E.D. boundaries and names are determined by the Representation Act passed by the Federal Government.

Size: The size of an E.D. is determined by the population, but, in general, urban E.D.'s have a larger population and smaller area than do mainly rural E.D.'s.

Use: The E.D. is not used for statistical presentation but is only a political entity.

Stability: The Representation Act can be amended after each decennial census in support of the concept of representation by population.

County, District and Regional Municipality

Definition: An aggregation of municipalities which, like a municipality, is a provincially organized area with some governmental responsibility.

Location: The boundaries are defined by provincial statute.

Size: There are no real limitations on size of a county but all Ontario counties are under 3,000 square miles in area. The Districts are much larger.

Use: Much of the population and agricultural data is published on a county base and these units are generally small enough that research and planning can be based upon them.

Stability: There are no real requirements for stability over time, but as yet, there have been no boundary alterations which prevent preparation of time series statistics for any of Ontario's past or present counties.

Census Metropolitan Area

Definition: A Census Metropolitan Area is the main labour market area of a continuous built-up area having a population of 100,000 or more. They are areas created by Statistics Canada.

Location: C.M.A.'s are located around urban centres and are usually known by the name of the major centre.

Size: The C.M.A. corresponds to a zone of commuting where people could normally change their place of work without changing their residence. CMA's must have at least 100,000 population and may include municipalities up to 20 miles from the built-up area.

Use: The C.M.A. is basically a publication unit which enables statistics for a "meaningful" area to all be presented together. It is apparent that these statistics would be more informative than those for the cities alone.

Stability: No real effort is made to ensure stability over time, and as the built-up area grows so will the overall area of the C.M.A.

Census Agglomeration

Definition:	A census agglomeration is a built-up area around an urban centre having a population of at least 1,000 and having additional adjacent urban municipalities or built-up areas with a density of 1,000 or more per square mile. The total population must exceed 2,000.
Location:	C.A.'s are located only around smaller urban centres and are, in effect, small C.M.A.'s.
Size:	The size is determined by population constraints with a minimum of 2,000 persons. If the population exceeds 100,000 persons the area becomes a C.M.A.
Use:	This unit is being used for the first time in 1971 and is basically for publication purposes.
Stability:	The C.A.'s are composed of complete municipalities and will change if municipal boundaries change. Growth will also cause expansion of the area included.

Municipality

Definition: An incorporated area having some form of governmental system. This includes townships, towns, villages, and cities.

Location: The boundaries are determined by the act of incorporation passed in the provincial parliament.

Size: The size only has a bearing upon the type of incorporation according to the following population schedule.

(1) City	more than 25,000
(2) Town	not less than 2,000
(3) Village	not less than 500
(4) Township	not less than 1,000
(5) Improvement District	not less than 50

Use: The municipality is the smallest areal unit used for publication of statistics except for the major urban areas in which census tracts have been identified.

Stability: There may be changes at any time through annexations and boundary changes but these must be officially altered by parliamentary act.

Area Aggregate

Definition: A grouping of contiguous enumeration areas.

Location: A.A.'s are located in both rural and urban areas.

Size: The A.A.'s are to range between 4,000 and 6,000 population but no areal constraints are imposed.

Use: A.A.'s are to be primarily for the publication of statistics.

Stability: The A.A.'s are to be kept strictly permanent but 1971 is the first year that they have been available. It would be more reasonable to expect that in the future some of them may be fractured to accommodate population growth.

Census Tract or C.T.

Definition: Census Tracts are statistical units that have been established in 23 Canadian Cities.

Location: C.T.'s are established only in urban areas.

Size: No particular size is specified but an effort is made to keep them homogenous which generally limits their size. They are intended to be small enough in area to be of use for research and planning.

Use: C.T.'s are the primary unit for publication of urban data.

Stability: These statistical areas are designed with intentions of maintaining stability for time series analysis.

Census District

Definition: A census district is the area for which one census commissioner is responsible.

Location: Census Districts are established in both rural and urban areas.

Size: The size is determined by the size of the electoral areas for which the supervisor is responsible. In rural areas it usually includes more than one township, from 10 to 15 E.A.'s, in urban areas 24 to 40 EA's is more common.

Use: The Census District is a functional area for the collection of the census data and is not used in publication or dissemination of the census results.

Stability: The Census District may be changed for each census depending upon the stability of its composite E.A.'s.

Enumeration Area or E.A.

Definition:

An E.A. contains the area covered by one enumerator in the collection of the census data.

Location:

E.A.'s are established in both rural and urban areas.

Size:

E.A. size is determined by population and the physical area that one enumerator could be expected to cover. In the urban areas this is usually a few city blocks, in most rural areas usually 10 to 20 square miles. They are also classified by density of population according to the following schedule:

A	more than 1000 per sq. mi.
B	500 - 999 per sq. mi.
C	100 - 499 per squ. mi.
D	10 - 99 per sq. mi.
E	2 - 9 per sq. mi.
F	Less than 1 per sq. mi.

Use:

The E.A. is the smallest area for which data is maintained and is the basic "building block' of the census system. All other census areas can be formed from these smallest census data units.

Stability:

The E.A. may be changed for each census depending upon the population as measured in the previous census.

1971 CENSUS MAPS

Reference maps are available for the 1971 census in two basic types. One type is that which is published in the report volumes of the census. These several bulletins are as follows:

A). The *Reference Map Bulletin* consists of about 60 maps showing the boundaries of:

1). the provinces (with Census Divisions, municipalities cities, town and villages).

2). components of Census Metropolitan Areas.

3). Components of Census Agglomerations (new in 1971).

B). The *Census Tract Bulletin* presents maps and population figures by census tract.

C). The *Unorganized Township Bulletin*.

D). The *Census of Agriculture* reports contain maps showing census divisions and municipalities.

A second group are those maps which are more detailed and are available only upon request.

A). The *National Topographic 1:500,000* series of maps will show census divisions (counties) and subdivisions (municipalities). A second set at the same scale will show electoral districts and enumeration areas.

B). The Enumeration Series provides complete documentation of all statistical areas and is presented in 3 groups.

The Group I or Rural EA, set is presented on various scales of the National Topographic series and presents boundaries for all counties, Electoral Districts, municipalities and all Enumeration Areas that can be outlined (approx. 40%) at this scale.

The Group 2, or Census Tract EA, set covers those EA's that fall within treated areas and presents only the boundaries of the EA's.

The Group 3, or Urban EA, set covers those urban EA's which are not within tracted areas.

ONTARIO GEOGRAPHICAL CLASSIFICATION MANUAL
FOR
PRODUCTION OF STATISTICS

This manual was not originally prepared with the 1971 census as its only purpose. The basic purpose of the manual is to facilitate the conversion of data from one departmental geographical system to another. If the Census of Population is simply regarded as an additional geographical system, then the use of the manual in preparing special tabulations from the Census can be seen.

At present the manual contains 36 areal breakdowns of the province employed by different Ontario Government organizations. Once census data becomes available it is expected that statistics will be easily prepared for each of these organization's areas.

ORGANIZATION OF THE DATA FOR RETRIEVAL

Tony Wellman

ORGANIZATION OF THE DATA FOR RETRIEVAL

P.V.S.T. codes

Clearly the vast amount of data available from the population census requires to be organized in such a way that the user is enabled to identify in a systematic way all that is potentially available from the population census, that part of this which is in the preplanned tabulation programme (publications, summary tapes, microfilm), and to specify concisely his request.

The framework chosen is derived from two considerations. These are -

(1) The general structure of population statistics in so far as these are composed of -

- a) The population which is being considered P
- b) The variables within the population that are being examined V
- c) The space or geographical area S
- d) The time to which the data relates T

(2) The structure of the questionnaire.

By considering a request in terms of its P.V.S.T. factors, whether the data is available can be determined through the use of the retrieval documentation which is discussed later.

P terms

Whether a term is defined as a P term at the collection stage (a first order P term) depends on whether separate questions are asked of this particular category of person. That is, if there are variables (V terms) unique to this category of persons, then this category is defined as a P term.

Thus, P terms can be identified from the questionnaire, new P terms being created wherever we see the 'skip to question x'.

In other words, every V term relates to a specific P term, that P term being the total universe, or population for which that variable is collected.

In addition to these 'first order' P terms defined at the collection stage, there are 'second order' P terms such as 'number of families' (derived from 'relationship to head of household' question). Second order P terms designate population on which information will be supplied, which had not been defined as populations at the collection stage.

Relating to these second order P terms are second order V terms, eg. income of family, derived by adding up separate incomes of members of the family.

To illustrate P & V terms, consider a request for -

Workforce	P
By Sex	V
Age	V

Marital Status	V
By Municipality for Ontario	S
1971 Census	T

Workforce is a 1st order P term, questions 33 to [illegible] on the questionnaire are questions unique to it.

V terms (V (Q), V (R)

Sex, age and marital status are the variables (V terms) within the population (P) (workforce) which are required.

However, the request is not completely specified until we indicate the particular breakdowns required for the V terms e.g. (age breakdown). So that this can be specified, the V term is broken up into two components, a V(Q) and a V(R). The V(Q) is the generic component, i.e. the question that is asked e.g. what is your sex?, what is your marital status?

The V(R) is the response breakdown. Clearly. the number of response breakdowns possible is restrained by the phrasing of the question e.g. if the respondent were only asked - are you married or unmarried?, it would not be possible to obtain a breakdown for divorced persons. However, within the restraints imposed, there are, in many cases, more than one breakdown possible. (The case of Age being a good example).

The existence of V(R) code make it possible to check what breakdowns (e.g. of age, marital status, etc.) are available on the various media for cross classifications required.

S term

S is the space term, or geographical breakdown, e.g. Ontario by Municipalities, E.A.'s etc. (i.e. breaking into the geographical levels listed in the previous presentation). If one requires particular data on only a particular municipality, one must check simply whether that particular P.V. configuration is available by the S factor 'Ontario by Municipalities'.

T term

This is 1971

3rd order terms

These are generated from 1st or second order terms by mathematical operation, e.g. ratios, percentages, means, etc.

RETRIEVAL DOCUMENTS

This then is the way the data is organized within the reference framework to enable users to search for and specify the information they require.

The reference framework comprises 3 components. These are -

(i) A dictionary
(ii) A classification manual
(iii) A directory

the components are linked through a common mnemonic code system.

The retrieval documents are in the process of being developed at the present time. A few minor details are still to be settled with regard to their format. However, the following general observations will hold true.

Dictionary

The dictionary defines what is collected on the population census. It will be arranged like the yellow pages. The definitions will be in alphabetical order and the dictionary will give synonyms referring the reader to the appropriate listing.

CHART 'A'

REFERENCE FRAMEWORK

DOCUMENTATION & ORGANISATION

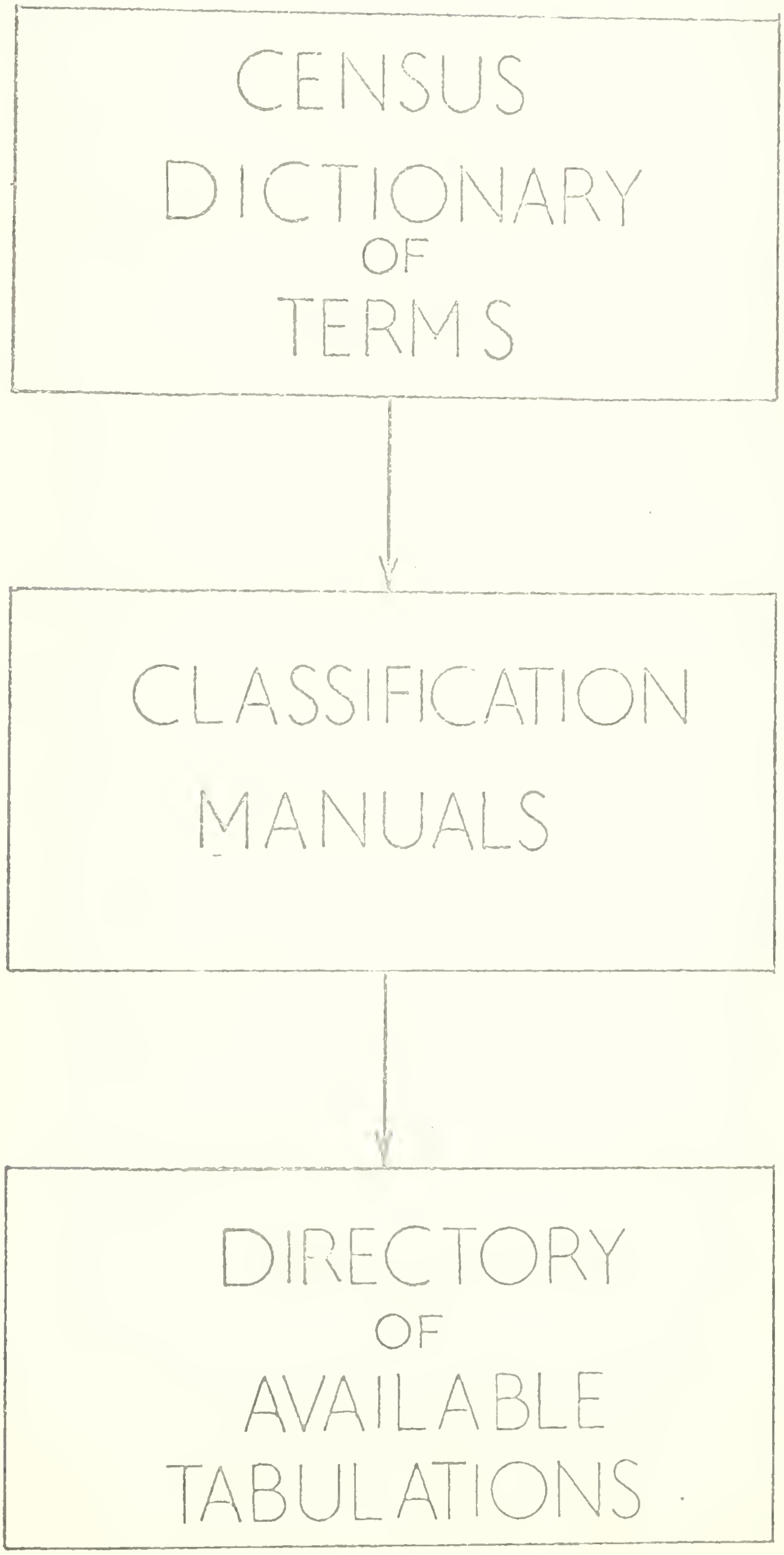

CENSUS DICTIONARY OF TERMS
CLASSIFICATION MANUALS
DIRECTORY OF AVAILABLE TABULATIONS

CHART 'B'

GENERAL REQUEST FORMULATION & SEARCH SEQUENCE

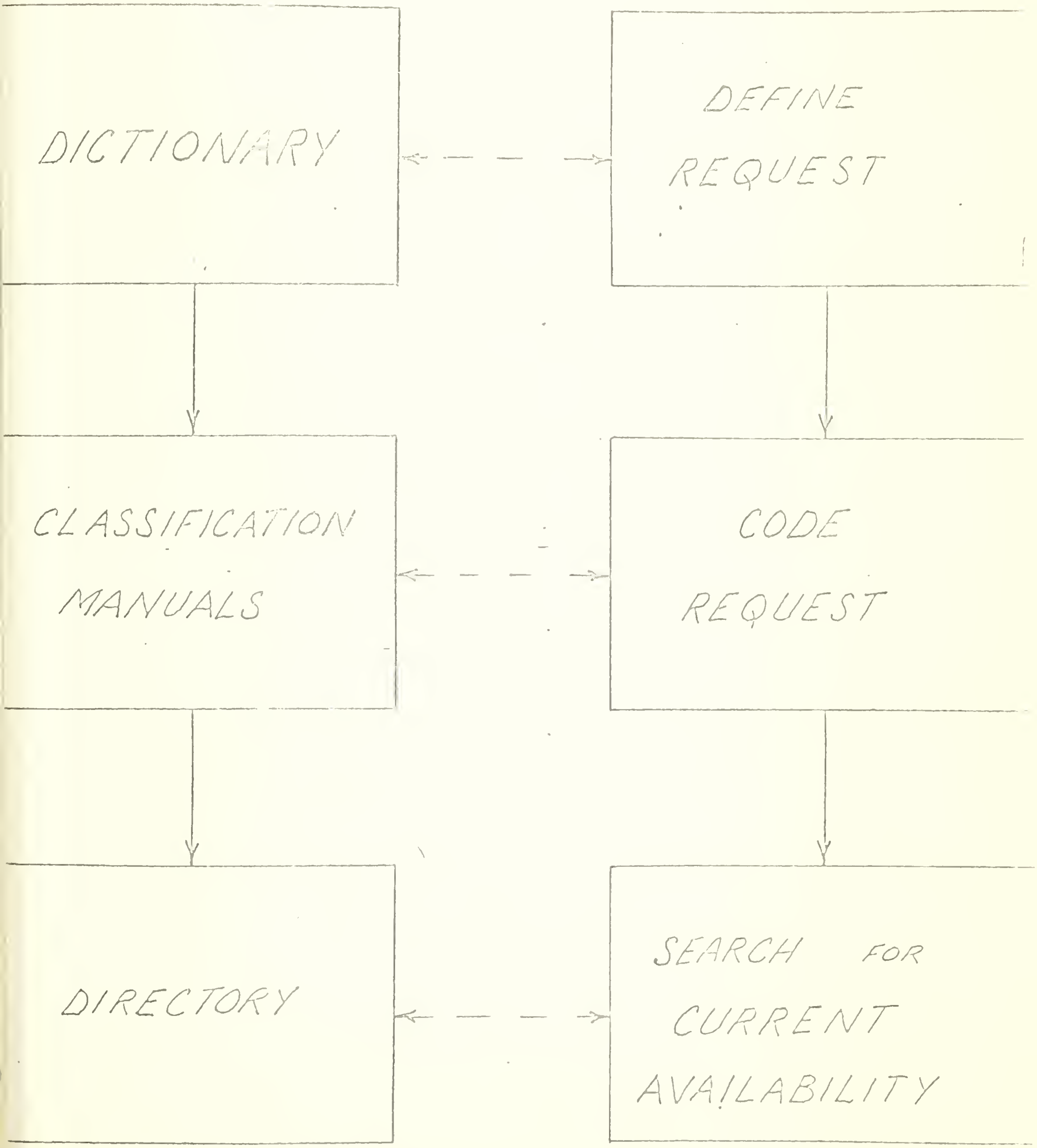


CHART 'C'

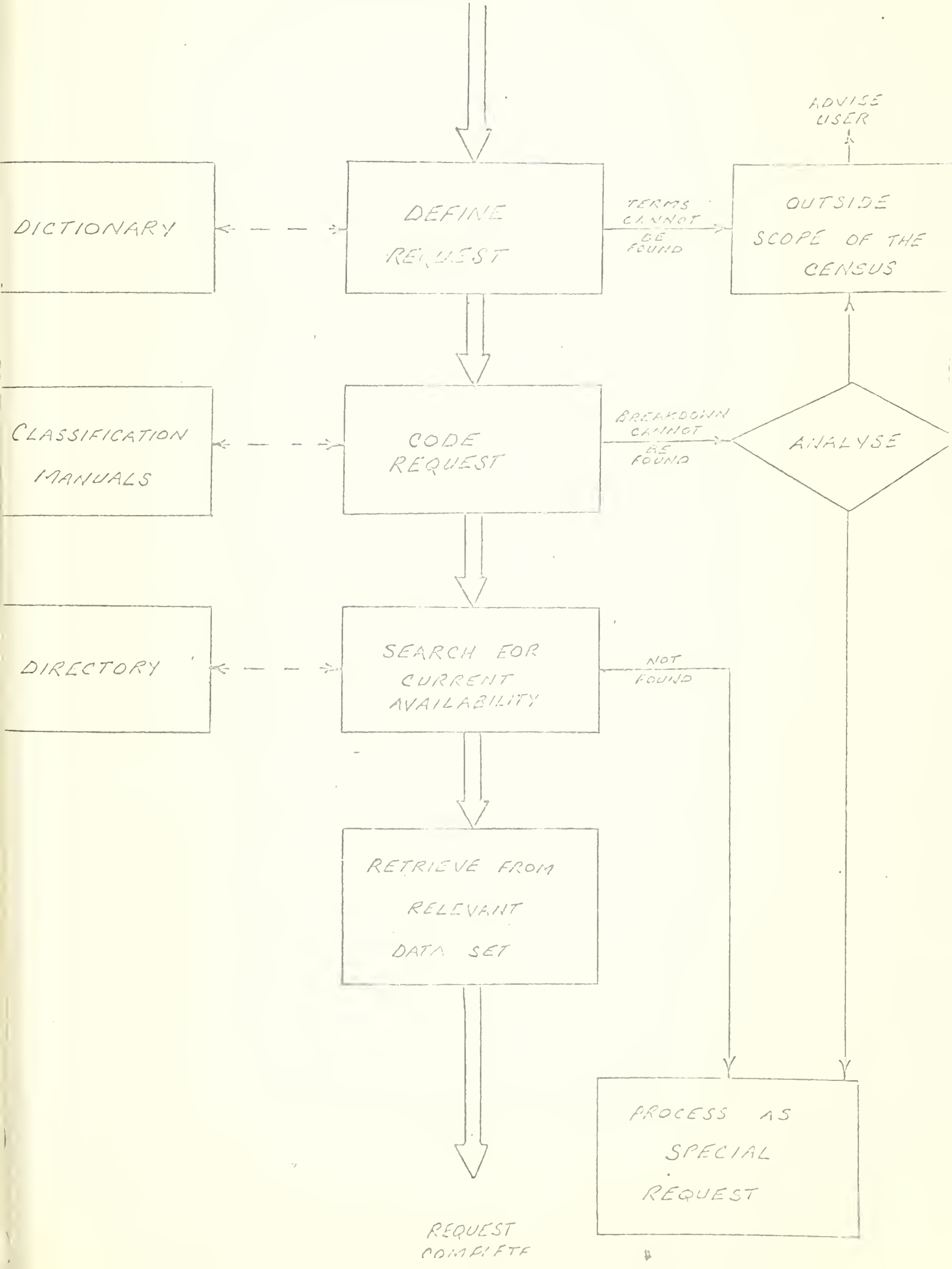

ADVISE USER
DICTIONARY
DEFINE REQUEST
TERMS CANNOT BE FOUND
OUTSIDE SCOPE OF THE CENSUS
CLASSIFICATION MANUALS
CODE REQUEST
BREAKDOWN CANNOT BE FOUND
ANALYSE
DIRECTORY
SEARCH FOR CURRENT AVAILABILITY
NOT FOUND
RETRIEVE FROM RELEVANT DATA SET
PROCESS AS SPECIAL REQUEST
REQUEST COMPLETE

CHART 'D'

DATA BASE

DATA FRAME

DEFINITION OF TERMS

- DICTIONARY

DATA BASE REFERENCES

- DIRECTORY
- CATALOGS (OF PUBLICATIONS, MICROFILMS, SUMMARY TAPES, ETC)
- TABULATION LISTS
- SUMMARY TAPE MANUALS

REFERENCE CODES

- CLASSIFICATION MANUALS
- OFFICE CODING MANUALS

to illustrate -

Fertility, See Children Ever Born

Entry:

CHILDREN EVER BORN .... 900 (CEB)

Refers to the number of children born alive, whether born of the present marriage or any previous marriage.

CENSUSES: 1971 (1/3 sample), 1961 (1/5 sample)

ASKED OF: all women in sample households who were 15 years old and older and who reported themselves as having been married at one time. (Includes those who reported themselves widowed, divorced or separated).

RESPONSES: 0, 1, 2 ... , 13, 14 and over

(Respondents were instructed to include children who died after birth as well as those residing elsewhere at Census time and to exclude adopted and step-children).

Using the mnemonic codes given in the dictionary, the request can be coded in terms of P.V(R) S.T.

Classification Manual

To determine the particular V(Q) breakdowns corresponding to each V(R), the Classification manual is referred to. This gives a code number and a descriptor for each V(Q) by its corresponding V(R).

Directory

Having coded the request in terms of P.V.G.T., the availability of the cross classification can be determined through the directory.

The directory will be ordered alphabetically by geographical level, i.e. there will be separate directories for each geographical level.

There will also be a directory listed in alphabetical order by V code and P code.

An idea of the type of format envisaged for the directory is shown on chart E. The numbers below the V(Q) terms give the V(R) breakdown available. To decode these numbers one would have to go to the classification manual.

Listed along side the code would be the media on which the data exists, e.g. publications, summary tapes, microfilm. If the tabulation is not listed, then it is a 'special request' which O.S.C. will forward to Ottawa to be taken off the appropriate file.

Where information is required on a population which is not a P term in the context of the census, e.g. divorced persons by age and occupation, a check in the dictionary will reveal that divorced persons is a response to the question on marital status which is asked of the total population. Thus, to check whether the information is available

one must check in the directory for the existence of a tabulation 'Total population by marital status, by age and occupation'. The information on divorced persons will be available within this tabulation.

User Request Form

Requests will be coded on the user request form. This is currently being developed by Statistics Canada. A standard request form will be used in all provinces.

Functions of this are -

(1) For user to specify requests.

(2) Exchange between provinces and between provinces and Statistics Canada to inform what information has been created on microfilm (and is thus available) while satisfying user requests.

(3) To keep a record of requests made in terms of usage of census data.

(4) As integral part of our indexing system.

CHART 'E'

DIRECTORY (e.g. by CENSUS DIVISION)

SEX	AGE - MOBILE		POPULATION	MEDIA
SEX 110	AGE 205		TPOP	
SEX 110	AGE 207		TPOP	
SEX 110	AGE 205		SCHPOP	
SEX 110	AGE 215	M.S. 012	POP15	
SEX 110	AGE 215	M.S. 010	POP15	
SEX 110	ETHNIC 405		TPOP	
SEX 110	ETHNIC 404		TPOP	

SOFTWARE FOR CENSUS RETRIEVAL

Bill Gray

SOFTWARE FOR CENSUS RETRIEVAL

The discussion will focus on the following:

(1) Program Packages and Systems considered for census retrieval

(2) Why Casper?
- efficiency
- drawbacks
- strong points

(3) How does Casper work?
- table building
- manipulation
- report writing

(4) Future possibilities
- statistical analysis
- terminal retrieval

The following notes are included for background.

Introduction to Some Statistics Canada Retrieval Programs

The packages considered are STATPAK, STATAPE and [illegible].

Both STATPAK and STATAPE employ a specification and retrieval language, very close to free-form plain language, called [illegible], and including automatic linkage through a variable dictionary.

STATPAK was specifically developed for the geocoded data base which is maintained on direct-access disc-packs with the data grouped into strings variable by variable. Retrieval is made for the user-selected study area from specific variable strings. File management involves the off-line selection of disc-packs according to the specific variables and areas. The package is included in this discussion because of the features included. Note should be made, of course, that retrievals are made with the geocoded census data in the individual questionnaire mode and, for this database will be operated extensively within Statistics Canada for special and small area requests.

STATAPE is similar to STATPAK but operates from magnetic tape data files. Analysis and tabulation is carried out on a cross-tabulation matrix basis in both packages; STATAPE is therefore of little application to the User Summary tapes.

Data is retrieved from the Census Master File and summed onto intermediate files (Production Summary Tapes) from which the tabulations are then produced.

A symbolic language is employed for specifications but no dictionary or prior set-up is used: the request must therefore be translated into field positions and lengths, etc.

For use with the summary tapes, basically the field listing and record selection features only are relevant: the user summary tapes being regarded as already summed.

Further documentation should be consulted for details on each of these packages.

Note regarding the User Summary Tapes:

The data files available on User Summary Tapes consist of counts and characteristics aggregated by various areas: they are in listings and cross tabulations of variables stored on magnetic tape.

As a result, the major package recommended to any summary tape user is one with a simple listing or report generator employing a selective retrieval language.

SOFTWARE PACKAGES

Name	Author	Description	System	Reference
CASPER	Statistics Canada	Cross-tabulation, analysis and report generator	S/360 OS-MVT 80K	Statistics Canada
STATAPE	Statistics Canada	Cross-tabulation and Statistical analysis	S/360 min. 120K-OS	Statistics Canada
STATPAK	Statistics Canada	Cross-tabulation and Statistical analysis (uses direct access files, record selection by geographic coordinates)	S/360 min. 120K-OS	Statistics Canada

CASPER DATA RETRIEVAL PROGRAM

INTRODUCTION

The objective of this presentation is to show you generally how the CASPER system works and what are some of its features.

What is CASPER?

CASPER is a generalised computer system written by Statistics Canada for the purpose of creating tabular reports.

The word CASPER is an abbreviation of Cross-classification And Special Purpose Extended Report-writer.

The system is user oriented, but slanted to people with data processing experience. With a little practice anyone can become proficient in the use of CASPER. Once you have reached a level of proficiency you have the use of a very powerful and economic system. The system gives you complete flexibility on just how your report is to be printed, and CASPER will permit you to print your tables in photocopying format so that your computer output can go directly to the print shop.

The CASPER system is now on a S/360-O.S., running under MVT, and HASP, with slight modification it can be adapted to run in MFT environment.

To implement the CASPER system the user codes control statements. The input to CASPER may be on key-punch cards, or an already existing file on magnetic tape or disk pack.

CASPER'S Major components

CASPER is made up of 6 major components that can be used independently or in conjunction with one another. If you are using more than one component, the output of one component becomes the input to the next component. The 6 components and their functions are:

.SUM360 will extract the data required from the input file to create a table. It will select only the data that is required and ignore the rest of the information on the file. It sets up this information in matrix form.

.MICMAN will manipulate the data within each table. It can perform mathematical calculations, or resequence the data in the table.

.MACMAN will manipulate the tables themselves. Its main use is to create total tables; it can also ressequence, select or delete specified tables.

.XPRINT will print the tables and can also calculate averages and medians if you require them. XPRINT is usually the last component coded in the CASPER system. It should however be followed by one or more XPRINT component if you require them.

The next 2 components were developed at a later date and are the result of additional requirements expressed by the users.

.FORMAT can invert your matrix. You might need the data in the table placed in a particular manner to permit you to perform certain calculations. However, when printing the tables it may be more feasible to present them in a different order. FORMAT can also calculate averages and medians if they are required.

.PHOTO like XPRINT will print your tables and calculate averages and medians, PHOTO however has a few more options. It permits you to code a table that has a 2 page spread in one pass, to manipulate on one level of data before printing the tables, prints your tables in photocopying format, and its coding is more compatible with the other components of CASPER than XPRINT.

DISSEMINATION OF DATA AT THE PROVINCIAL LEVEL

Tony Wellman

DISSEMINATION OF DATA AT THE PROVINCIAL LEVEL

O.S.C. will be acting as a focal point for receiving and meeting requests for census data from the provincial departments.

As referred to earlier, for this role we shall be equipped with the three media -

Publications
Summary tapes
Microfilm

and access through Statistics Canada to -

Master File
Production Summary Tapes
Geocoded data base

The provincial departments will hold copies of the retrieval documents already mentioned:

User Request Forms
Dictionary
Classification Manual
Directory

Requests will be submitted on the user request form.

O.S.C. will check whether it is possible to retrieve the data from the media available here.

Where the request is satisfied from the summary tapes, the computer print-out will be microfilmed and added to the microfilm library for ready access in case a similar request is received from another department.

Where the request is a special request, i.e. one that cannot be met from the media within O.S.C., then O.S.C. will submit it directly to Ottawa for processing. It will then be returned to O.S.C. who will submit it to the user and retain a copy to be microfilmed for addition to the library.

To enable users to fully utilize data from the census, workshops will be arranged on specialist areas according to demand.

O.S.C. will also upon demand and as resources permit, act in a consulting capacity for statistical analysis.

HA/747.65/.O93/1972
Ontario. Dept. of Treasury
Overview of 1971
census of population gkuw
c.1 tor mai